Current Sample Size Practices in the Psychometric Evaluation of Patient-Reported Outcome Measures for Use in Clinical Trials

Theresa Coles, Wen-Hung Chen, Lauren Nelson, Valerie Williams, Nicole Williams, Lori McLeod

RTI Health Solutions, Research Triangle Park, NC, United States

BACKGROUND

- **The situation:** Sample size affects the robustness of psychometric results.
- **Why is this a problem?**
  - Clinical trials are powered for the statistical methods used to evaluate the primary endpoint, which may or may not include a patient-reported outcome (PRO) measure.
  - Currently, there is little guidance for psychometric sample size requirements for the development of PRO measures for use in clinical trials.
- **What are the consequences?**
  - Sample sizes that are too large or too small may influence conclusions regarding the reliability and validity of the PRO measures used in clinical trials.
    - With sample sizes that are too large, the evaluation may be overpowered for the PRO evaluation, resulting in more type I error (i.e., detection of an effect that is not present).
    - With sample sizes that are too small, the applicable analysis methods may be limited, models may not converge, and results will not be robust; insufficient power results in type II error (i.e., failure to detect an effect that is present).
- **First step to change:** The first step toward encouraging a more rigorous approach to PRO sample size practices is to describe current sample size practices in the literature.

OBJECTIVE

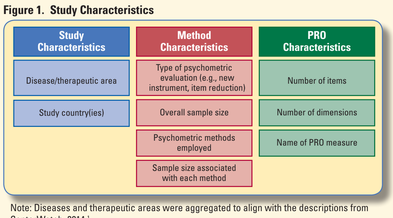
- To describe the statistical methods and corresponding sample size practices in published psychometric evaluation studies of PRO measures likely developed for use in clinical trials over the past decade.

METHODS

Systematic Literature Review and Data Extraction

- Data source: PubMed
- Eligibility criteria
  - Articles published in peer-reviewed journals
  - Validation or psychometric evaluation articles (screeners and preference questionnaires were excluded because they require a unique set of evaluation methods)
  - PRO measure likely developed for use in clinical trials
  - No review articles, opinion pieces, duplicate articles, or gray literature
  - Articles written in English
  - Articles published on or after January 1, 2004
- Literature review and data-extraction process
  1. The search strategy (available on request) consisted of four concept blocks designed to capture mentions of (1) PROs and (2) clinical trials and (3) psychometric and (4) development/validation studies in PubMed. The search was performed on May 15, 2014.
  2. The primary reviewer (TC) trained a team of four reviewers (WHC, LN, VW, LM) and provided a set of instructions and a template for extracting data.
  3. Five reviewers were assigned titles and abstracts for review.
  4. Reviewers independently read titles and abstracts for relevance. Abstracts that met the eligibility criteria were included in the analysis.
     - If an abstract did not mention a sample size, the full-text article was obtained for review.
  5. The study characteristics presented in Figure 1 were tabulated for each article.
  6. Reviewers recorded the psychometric methods used in each article, as well as the sample size associated with each method. Figure 2 lists the psychometric methods evaluated in this study by the complexity of the analysis methods.

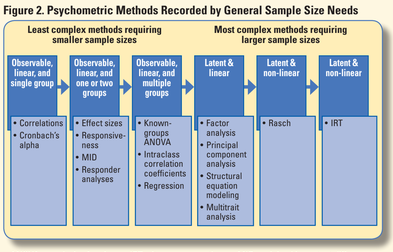
Figure 1. Study Characteristics



Note: Diseases and therapeutic areas were aggregated to align with the descriptions from CenterWatch, 2014.[1]

Note: The number of items and dimensions were not included in a large proportion of abstracts (or full-text papers where applicable). Almost 43% did not report the number of items (42.5%) and 54.1% did not report the number of dimensions. Therefore, these results are not presented.

Figure 2. Psychometric Methods Recorded by General Sample Size Needs



ANOVA = analysis of variance; IRT = item response theory; MID = minimal important difference.

Note: Sample size requirements vary depending on the complexity of the PRO instrument and psychometric method.

Data-Collection Conventions

- If more than one sample size was reported in an abstract because a measure was evaluated more than once, the smallest sample size was reported because it represented the lowest common denominator.
- If countries were not mentioned in the abstract, the corresponding author's location was used as a proxy.
- Quality-control procedures were as follows:
  - An independent quality-control reviewer (NW) confirmed the results by examining a random 10% of articles excluded from the analyses and another random 10% of articles included in the analyses.
  - Discrepancies were resolved based on consensus between the abstract reviewer(s) and the independent quality-control reviewer.

Analytic Methods

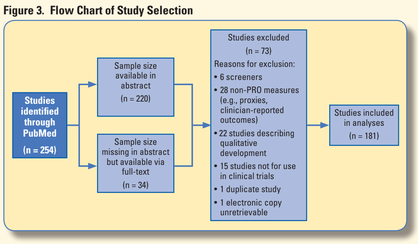
- Analyses were descriptive in nature. No statistical tests were performed.

RESULTS

Study Selection

Figure 3 illustrates the results of the study selection process.

Figure 3. Flow Chart of Study Selection



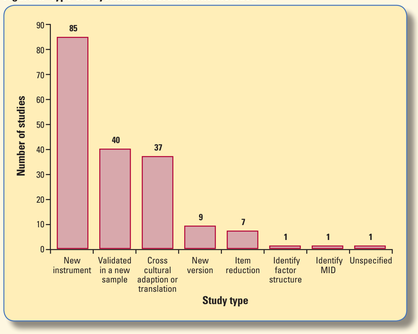
Quality Control

- Results: 7 excluded abstracts were reviewed and 18 included studies (abstracts or full-text articles as applicable) were reviewed; 3.4% of entries did not match and were resolved. Given the low error rate, further verification was not warranted.

Description of Studies

- Twenty-six countries were represented in the studies included in these analyses. Studies were most often based in the United States (36.5%), United Kingdom (13.3%), and Germany (6.1%). Approximately 9% of studies were conducted in multiple countries.
- Among the 181 studies included in the final analyses, the most frequently studied therapeutic areas were oncology (13.8%), gastroenterology (11.6%), neurology (10.5%), psychiatry/psychology (8.3%), and urology (6.6%). Overall, 26 therapeutic areas were represented in this study.
- Figure 4 presents a description of study types included in these analyses.

Figure 4. Type of Psychometric Evaluations Included



Sample Size Practices

- Overall study sample sizes ranged from 39 to 14,038. The overall mean sample size was 527.7 (standard deviation [SD], 1241.38; median, 249.0; minimum, 39; maximum, 14,038).
- Table 1 presents the number (and percentage) of studies that reported the use of each psychometric method. The most frequently reported psychometric method was Cronbach's alpha to quantify internal consistency (n = 157, 86.7%).
- Within each study, sample sizes were recorded by method where available. Table 2 presents descriptive statistics for the samples employed by method. Figure 5 illustrates the distribution of sample sizes used for each method. To show detailed distributions, the maximum sample size displayed in the figure was set to 2,000, which represents approximately 97% of the sample.
- The mean sample size by method ranged between 280.2 (median, 185.0) and 1,001.9 (median, 225.0).
- The minimum and maximum values are of some concern, as studies could be underpowered or overpowered depending on the complexity of the measure or method.

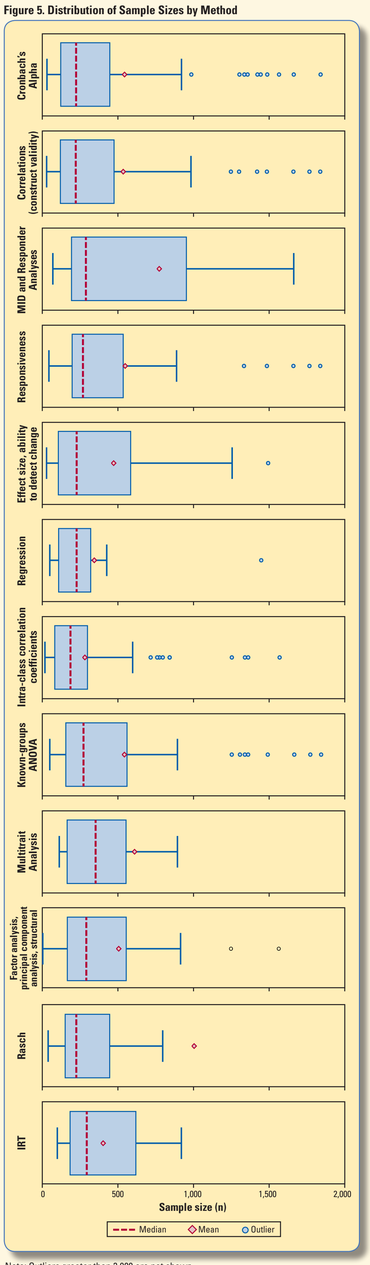
Table 1. Psychometric Methods Employed in Studies

Psychometric Method	Studies Reported Using Method n (%)
Cronbach's alpha	157 (86.7)
Correlations (construct validity)	136 (75.1)
MID, responder analyses	12 (6.6)
Responsiveness	50 (27.6)
Effect size	25 (13.8)
Regression	9 (5.0)
Intraclass correlation coefficients	81 (44.8)
Known-groups ANOVA	83 (45.9)
Multitrait analysis	15 (8.3)
Factor analysis, principal component analysis, structural equation modeling	79 (43.7)
Rasch	19 (10.5)
IRT	4 (2.2)

Table 2. Sample Size Descriptive Statistics by Method

Psychometric Method	Number of Studies	Mean Sample Size Used in Study	SD	Median	Minimum	Maximum
Cronbach's alpha	157	545.8	1324.6	227.0	34	14,038
Correlations (construct validity)	136	538.5	1347.6	225.5	34	14,038
MID, responder analyses	12	775.4	1,119.4	290.5	74	4,000
Responsiveness	50	551.9	742.3	273	49	4,000
Effect size	25	468.6	598.2	225.0	26	2,674
Regression	9	342.0	431.5	227.0	50	1,443
Intra-class correlation coefficients	81	280.2	319.7	185.0	17	1565.0
Known-groups ANOVA	83	542.5	829.6	271.0	50	5,521
Multitrait analysis	15	609.5	963.4	351.0	112	4,000
Factor analysis, principal component analysis, structural equation modeling	79	509.7	795.7	295.0	8	5,521
Rasch	19	1,001.9	3,163.3	225.0	39	14,038
IRT	4	400.5	356.1	293.0	100	916

Figure 5. Distribution of Sample Sizes by Method



Note: Outliers greater than 2,000 are not shown.

DISCUSSION

- Ideally, researchers should consider the complexity of the PRO measure, its intended use, and the purpose of the evaluation when deciding on a study sample size.
- This study provided an opportunity to review the current PRO evaluation sample size practices in the published literature using a systematic literature review. To our knowledge, it is the first study to describe these practices in the literature.
- Sample sizes employed for the psychometric evaluation of PROs developed for use in clinical trials varied widely, overall and by method.
  - Descriptive statistics indicate that sample sizes were inconsistently selected by the complexity of the psychometric method.
- Researchers rarely provide a rationale for sample sizes used in most clinical trial–related psychometric studies.
- Key limitations of this study are as follows:
  - PRO measures included in this study were not compared with clinical trial information to verify their use in clinical trials.
  - Abstracts are designed to be brief, and sample size information for each method was not always noted. Full-text articles were not always clear about the sample sizes employed for each method.

CONCLUSION

- Additional studies should work toward developing best practices for PRO sample size guidelines in clinical trials.

REFERENCE

1. CenterWatch. Medical Therapeutic Area Descriptions. Available at: http://www.centerwatch.com/clinical-trials/listings/therapeutic-description.aspx. Accessed October 16, 2014.

CONTACT INFORMATION

Theresa Coles, MS
Senior Health Outcomes Analyst

RTI Health Solutions
200 Park Offices Drive
Research Triangle Park, NC 27709

Phone: +1.919.316.3843
E-mail: tcoles@rti.org

Presented at:
ISPOR 17th Annual European Congress
November 8-12, 2014
Amsterdam, The Netherlands